Stochastic Processes and their Applications 36 (1990) 353-370
North-Holland

EFFICIENCY OF ESTIMATORS FOR PARTIALLY SPECIFIED FILTERED MODELS

P.E. GREENWOOD*

Department of Mathematics, University of British Columbia, Vancouver, B.C., Canada V6T 1Y4

W. WEFELMEYER**

Department of Mathematics, University of Cologne, 5000 Cologne 41, West Germany

Received 9 May 1989
Revised 23 November 1989

Let $X_{n1}, \ldots, X_{nn}$ be counting processes and let $Y_{n1}, \ldots, Y_{nn}$ be vector-valued covariate processes. Assume that the intensity processes of the X_{ni} with respect to the filtration generated by X_{ni} and Y_{ni} are known up to a (possibly infinite-dimensional) parameter, but that the distribution of X_{ni} and Y_{ni} is unspecified otherwise. We give conditions under which the partially specified likelihood in the sense of Gill-Slud-Jacod is locally asymptotically normal. We show that the partially specified likelihood determines a covariance bound in the sense of a Hájek-LeCam convolution theorem for estimating functionals of the underlying parameter. The theorem shows that the Huffer-McKeague estimator is efficient in Aalen's additive risk model, and that the Cox estimator for the regression coefficients and a Breslow-type estimator for the integrated baseline hazard are efficient in Cox's and in Prentice and Self's proportional hazards models.

AMS 1980 Subject Classification. 62G05, 62G20, 62M09.

Efficient nonparametric estimation * partially specified model * partially specified likelihood * proportional hazards model * additive risk model

1. Introduction

Many models in survival analysis can be conveniently described as follows. Let $X_{n1}, \ldots, X_{nn}$ be counting processes, and let $Y_{n1}, \ldots, Y_{nn}$ be vector-valued covariate processes over a time interval $[0, 1]$. Assume that no two counting processes jump simultaneously. Let $\mathbb{F}_n = (\mathscr{F}_{nt})_{t\in[0,1]}$ denote the filtration generated by $X_n = (X_{n1}, \ldots, X_{nn})$, and $\mathbb{G}_n = (\mathscr{G}_{nt})_{t\in[0,1]}$ the filtration generated by X_n and $Y_n = (Y_{n1}, \ldots, Y_{nn})$ together. For ϑ in some set Θ let $a_{n\vartheta} = (a_{n1\vartheta}, \ldots, a_{nn\vartheta})$ be a predictable process. We assume that the martingale problem associated with X_n and $a_{n\vartheta}$ has a solution: There exists a probability measure such that each X_{ni} admits a

* Work supported by an NSF Visiting Professorship and NSERC, Canada.
** Work supported by a Feodor Lynen grant of the Alexander von Humboldt foundation, and NSERC, Canada.

Doob–Meyer decomposition with respect to $\mathbb{G}_n$ and this probability measure, of the form

$$X_{ni}(t) = M_{ni\vartheta}(t) + \int_0^t a_{ni\vartheta}(s)\,\mathrm{d}s, \quad t \in [0,1]. \tag{1.1}$$

In other words: we assume that the intensity process $a_{ni\vartheta}$ of X_{ni} is known up to a (possibly infinite-dimensional) parameter $\vartheta \in \Theta$. We are interested in efficiently estimating ϑ, or a functional of it.

Let $\mathscr{P}_{n\vartheta}$ denote the family of all solutions of the martingale problem associated with X_n and $a_{n\vartheta}$.

If $\mathbb{G}_n = \mathbb{F}_n$, there is no additional information about ϑ in the covariates Y_{ni}. Hence it suffices to observe the counting processes X_{ni}. Their distribution is determined by the intensity processes $a_{ni\vartheta}$. This means that the model is *fully specified.* The family $\mathscr{P}_{n\vartheta}$ reduces to a single probability measure $P_{n\vartheta}$. The model can be studied using the following representation of the log-likelihood between τ and ϑ:

$$\begin{aligned}&\log(\mathrm{d}P_{n\tau}|\mathscr{F}_{n1}/\mathrm{d}P_{n\vartheta}|\mathscr{F}_{n1})\\ &\quad = \sum_{i=1}^n \left(\int_0^1 \log(a_{ni\tau}(s)/a_{ni\vartheta}(s))\,\mathrm{d}X_{ni}(s) - \int_0^1 (a_{ni\tau}(s) - a_{ni\vartheta}(s))\,\mathrm{d}s \right).\end{aligned} \tag{1.2}$$

(Set $\log 0 = -\infty$.) For multivariate point processes, this representation was introduced by Jacod (1975, p. 250, Theorem (5.1)); see also Jacod and Shiryaev (1987, p. 190, Theorem 5.45). A specialization to multivariate counting processes (no two of which jump simultaneously) is given in Dzhaparidze (1985). He also studies efficiency in this case.

We are interested in the case where $\mathbb{G}_n$ is larger than $\mathbb{F}_n$. Then the joint distribution of X_{ni} and Y_{ni} is not determined by the intensity process $a_{ni\vartheta}$. This means that the model is *partially specified* in the sense of Greenwood (1988). In particular, the likelihoods are unspecified. Hence there is no unambiguous efficiency concept. One can, however, still write down the right-hand expression in (1.2) and define the *partially specified log-likelihood* between τ and ϑ as

$$L_{n\vartheta\tau} = \sum_{i=1}^n \left(\int_0^1 \log(a_{ni\tau}(s)/a_{ni\vartheta}(s))\,\mathrm{d}X_{ni}(s) - \int_0^1 (a_{ni\tau}(s) - a_{ni\vartheta}(s))\,\mathrm{d}s \right). \tag{1.3}$$

For the counting process setting considered here, the partially specified likelihood was introduced by Gill (1985) and Slud (1986, Chapter 6), generalizing the discrete-time version proposed by Cox (1975). A corresponding concept for general semimartingales was introduced and studied by Jacod (1987, 1990a).

An efficiency concept for the partially specified model can now be introduced as follows. Fix $\vartheta \in \Theta$ and $P_n \in \mathscr{P}_{n\vartheta}$. For $\tau \in \Theta$ define $Q_{n\tau}$ by

$$\mathrm{d}Q_{n\tau} = \exp(L_{n\vartheta\tau})\,\mathrm{d}P_n.$$

Then $L_{n\vartheta\tau}$ is the log-likelihood between $Q_{n\tau}$ and $Q_{n\vartheta} = P_n$, and we can base an efficiency concept on $L_{n\vartheta\tau}$.

The paper is organized as follows. In Section 2 we give conditions under which the model $\{Q_{n\tau}: \tau \in \Theta\}$ fulfills an infinite-dimensional version of local asymptotic normality, and we formulate a convolution theorem for finite-dimensional functionals of the parameter. Applications are considered in Sections 3 and 4.

In Section 3 we treat the *additive risk model* with intensity process

$$a_{ni\vartheta}(s) = Y_{ni}(s)\lambda(s). \tag{1.4}$$

Here Y_{ni} is a row vector of *covariate processes* and λ is a (column) vector of *hazard functions*. The parameter is λ. This model was introduced by Aalen (1980). He suggested a least squares estimator for the vector of *cumulative hazards*

$$\int_0^t \lambda(s)\,\mathrm{d}s.$$

Huffer and McKeague (1987) determined its asymptotic covariance and suggested a *weighted* least squares estimator. The asymptotic covariance of the latter was obtained in McKeague (1988b); it turns out to be smaller than Aalen's. (An analogous result holds for grouped data; see McKeague, 1988a.) A heuristic explanation for the choice of weights in the Huffer–McKeague estimator is given in McKeague and Utikal (1990). We show that the Huffer–McKeague estimator is efficient.

In Section 4 we study the *proportional hazards model* with intensity process

$$a_{ni\vartheta}(s) = C_{ni}(s)\lambda(s)\exp(\beta' Y_{ni}(s)). \tag{1.5}$$

Here C_{ni} is a *censoring process*, taking only values 0 and 1, Y_{ni} is a vector of *covariate processes*, β is a vector of *regression coefficients*, and λ is the *baseline hazard function*. This generalization of the Cox (1972) model from i.i.d. random variables to counting processes and time-dependent covariates is due to Andersen and Gill (1982). The parameter is $\vartheta = (\beta, \lambda)$. Andersen and Gill determine the asymptotic distribution of Cox's (1972) maximum partial likelihood estimator for the regression coefficients β, and of a version of Breslow's (1972, 1974) estimator for the *cumulative baseline hazard*

$$\int_0^t \lambda(s)\,\mathrm{d}s.$$

We show that these estimators are efficient.

For a different treatment of efficiency in the fully specified versions of the additive risk and the proportional hazards model see Andersen et al. (1990, Chapter 8).

Prentice and Self (1983) suggest replacing the exponential function in Cox's proportional hazards model (1.5) by some other *known* (smooth) function. Following Andersen and Gill (1982), they determine the asymptotic distributions of appropriate variants of Cox's and Breslow's estimators. We indicate briefly that these variants are also efficient.

2. A convolution theorem

For $n \in \mathbb{N}$ let $(\Omega_n, \mathscr{F}_n)$ be a measurable space. Let

$$X_{n1}(t), \ldots, X_{nn}(t), \quad t \in [0, 1],$$

be counting processes. Let

$$Y_{n1}(t), \ldots, Y_{nn}(t), \quad t \in [0, 1],$$

be p-dimensional vectors of covariate processes. Set

$$X_n = (X_{n1}, \ldots, X_{nn}), \qquad Y_n = (Y_{n1}, \ldots, Y_{nn}).$$

Let $\mathbb{F}_n = (\mathscr{F}_{nt})_{t \in [0,1]}$ denote the filtration generated by X_n, and let $\mathbb{G}_n = (\mathscr{G}_{nt})_{t \in [0,1]}$ denote the filtration generated by X_n and Y_n together. Let Θ be an arbitrary parameter space. For $\vartheta \in \Theta$ let $a_{n\vartheta} = (a_{n1\vartheta}, \ldots, a_{nn\vartheta})$ be a $\mathbb{G}_n$-predictable process. Assume that the family $\mathscr{P}_{n\vartheta}$ of all probability measures P_n on $\mathscr{F}_n$ with the following properties is not empty. The components of X_n are P_n-a.s. finite, and no two of them jump simultaneously. The Doob–Meyer decomposition of X_{ni} with respect to $\mathbb{G}_n$ and P_n is

$$X_{ni}(t) = M_{ni\vartheta}(t) + A_{ni\vartheta}(t), \quad t \in [0, 1],$$

with $M_{ni\vartheta}$ a martingale and $A_{ni\vartheta}$ the compensator, and

$$A_{ni\vartheta}(t) = \int_0^t a_{ni\vartheta}(s)\, ds, \quad t \in [0, 1].$$

The predictable density $a_{ni\vartheta}$ is called the *intensity process.*

Fix $\vartheta \in \Theta$ and $P_n \in \mathscr{P}_{n\vartheta}$. Let V be a linear space, $(\cdot, \cdot)$ an inner product on V, and $\|\cdot\|$ the corresponding norm. For $i = 1, \ldots, n$ let

$$D_{ni} : V \times \Omega_n \times [0, 1] \to \mathbb{R}$$

be linear in the first variable, $v \in V$, and adapted to $\mathbb{G}_n$ for fixed v. Choose a *rate* $c_n \to \infty$. Assume that for $v \in V$ there exists a sequence (ϑ_{nv}) in Θ with the following properties:

(1) The intensity processes $a_{ni\vartheta}$ are *Hellinger differentiable* at ϑ with *derivative* D_{ni}:

$$\sum_{i=1}^n \int_0^1 ((a_{ni\vartheta_{nv}}(s)/a_{ni\vartheta}(s))^{1/2} - 1 - \tfrac{1}{2} c_n^{-1} D_{ni}(v)(s))^2 a_{ni\vartheta}(s)\, ds \to 0 \quad (P_n). \tag{2.1}$$

We assume, for convenience, that $a_{ni\vartheta}(s) = 0$ implies $a_{ni\vartheta_{nv}}(s) = 0$ for ds-a.e. $s \in [0, 1]$, P_n-a.s.

Proposition 2.5 remains true if we replace (2.1) by

$$\sum_{i=1}^n \int_0^1 (a_{ni\vartheta_{nv}}(s)/a_{ni\vartheta}(s) - 1 - c_n^{-1} D_{ni}(v)(s))^2 a_{ni\vartheta}(s)\, ds \to 0 \quad (P_n). \tag{2.1'}$$

Since (2.1) is tied to the Hellinger process, it is more natural and, in general, less restrictive than (2.1′). Under (2.2) and (2.3), relation (2.1′) implies (2.1). Condition (2.1′) is easier to check in the examples of Sections 3 and 4.

(2) The derivatives D_{ni} fulfill a *Lindeberg condition*: For $\varepsilon > 0$,

$$c_n^{-2} \sum_{i=1}^{n} \int_0^1 D_{ni}(v)(s)^2 1\{|D_{ni}(v)(s)| > \varepsilon c_n\} a_{ni\vartheta}(s)\, ds \to 0 \quad (P_n). \tag{2.2}$$

(3) The Hellinger limit is *nonrandom*:

$$c_n^{-2} \sum_{i=1}^{n} \int_0^1 D_{ni}(v)(s)^2 a_{ni\vartheta}(s)\, ds \to \|v\|^2 \quad (P_n). \tag{2.3}$$

It is condition (2.3) which makes the model locally asymptotically normal as opposed to 'mixed normal'.

Note that V, $(\cdot,\cdot)$, $\|\cdot\|$, D_{ni} and ϑ_{nv} depend on ϑ. The norm $\|v\|$ determines how difficult it is to distinguish, asymptotically, between ϑ and ϑ_{nv}. We call the corresponding inner product $(\cdot,\cdot)$ the *acuity*.

Conditions (2.1) to (2.3) are similar to Condition Φ introduced by Dzhaparidze (1985) for a different (fully specified) model. The following proposition shows that they imply an infinite-dimensional version of *local asymptotic normality* for the partially specified log-likelihood

$$L_{nv} = L_{n\vartheta\vartheta_{nv}}, \tag{2.4}$$

with $L_{n\vartheta\tau}$ defined in (1.3). The proof simplifies for models with multiplicative intensity; see Greenwood and Wefelmeyer (1989a).

2.5. Proposition. *If* (2.1) *to* (2.3) *hold, then*

$$L_{nv} = Z_n(v) - \tfrac{1}{2}\|v\|^2 + o_{P_n}(1), \tag{2.6}$$

with

$$Z_n(v) = c_n^{-1} \sum_{i=1}^{n} \int_0^1 D_{ni}(v)(s)\, dM_{ni\vartheta}(s), \tag{2.7}$$

and $Z_n(v)$ *is asymptotically normal under* P_n *with variance* $\|v\|^2$.

Proof. Note that $\log x = 2\log(1 + x^{1/2} - 1)$ and $x - 1 = 2(x^{1/2} - 1) + (x^{1/2} - 1)^2$. Define

$$r(x) = \log(1+x) - x + \tfrac{1}{2}x^2.$$

Set $a_{niv} = a_{ni\vartheta_{nv}}$ and suppress ϑ whenever convenient. By Taylor expansion, we can write the partially specified log-likelihood (2.4) as

$$L_{nv} = 2\sum \int ((a_{niv}/a_{ni})^{1/2} - 1)\, dM_{ni}(s)$$

$$- 2\sum \int ((a_{niv}/a_{ni})^{1/2} - 1)^2 a_{ni}\, ds + R_n, \tag{2.8}$$

where the remainder term is

$$R_n = 2\sum \int r((a_{niv}/a_{ni})^{1/2}-1)\,\mathrm{d}X_{ni}(s)$$

$$-\sum \int ((a_{niv}/a_{ni})^{1/2}-1)^2\,\mathrm{d}M_{ni}(s). \tag{2.9}$$

Below we show that $R_n = \mathrm{o}_{P_n}(1)$. To complete the proof, replace the expressions $(a_{niv}/a_{ni})^{1/2}-1$ in the first sum of (2.8) by $\frac{1}{2}c_n^{-1}D_{ni}(v)$, and replace the second sum by $\frac{1}{4}\|v\|^2$. This is justified by Hellinger differentiability (2.1) and condition (2.3). By the Lindeberg condition (2.2), the sequence $Z_n(v)$ is asymptotically normal under P_n with variance $\|v\|^2$; see e.g. Jacod and Shiryaev (1987, p. 429, Theorem 3.6b).

(i) We show that

$$\sum \int r((a_{niv}/a_{ni})^{1/2}-1)1\{|(a_{niv}/a_{ni})^{1/2}-1|>\varepsilon\}\,\mathrm{d}X_{ni}(s) = \mathrm{o}_{P_n}(1). \tag{2.10}$$

It suffices to prove that the probability that any X_{ni} jumps in $\{|(a_{niv}/a_{ni})^{1/2}-1|>\varepsilon\}$ is negligible:

$$P_n\left\{\sum \int 1\{|(a_{niv}/a_{ni})^{1/2}-1|>\varepsilon\}\,\mathrm{d}X_{ni}(s)\leq 1\right\}\to 0.$$

From $\sum \mathrm{d}\langle M_{ni}\rangle = \sum a_{ni}\,\mathrm{d}s$ and Lenglart's inequality (Jacod and Shiryaev, 1987, p. 35, Lemma 3.30) we see that it suffices to prove

$$\sum \int 1\{|a_{niv}/a_{ni})^{1/2}-1|>\varepsilon\}a_{ni}\,\mathrm{d}s = \mathrm{o}_{P_n}(1).$$

This follows from (2.1) to (2.3):

$$\sum \int 1\{|(a_{niv}/a_{ni})^{1/2}-1|>\varepsilon\}a_{ni}\,\mathrm{d}s$$

$$\leq \varepsilon^{-2}\sum \int ((a_{niv}/a_{ni})^{1/2}-1)^2 1\{|(a_{niv}/a_{ni})^{1/2}-1|>\varepsilon\}a_{ni}\,\mathrm{d}s$$

$$= \tfrac{1}{4}\varepsilon^{-2}c_n^{-2}\sum \int D_{ni}(v)^2 1\{|D_{ni}(v)|>\varepsilon c_n\}a_{ni}\,\mathrm{d}s + \mathrm{o}_{P_n}(1)$$

$$= \mathrm{o}_{P_n}(1).$$

(ii) Since (2.10) is of order $\mathrm{o}_{P_n}(1)$ for each $\varepsilon>0$, the same is true for some sequence $\varepsilon_n\to 0$. To show that the first term of the remainder (2.9) is of order $\mathrm{o}_{P_n}(1)$, it remains to prove

$$\sum \int r((a_{niv}/a_{ni})^{1/2}-1)1\{|(a_{niv}/a_{ni})^{1/2}-1|\leq\varepsilon_n\}\,\mathrm{d}X_{ni}(s) = \mathrm{o}_{P_n}(1).$$

Note that $r(x) \leqslant 2|x|^3$ for $|x| \leqslant \frac{1}{2}$. Hence, again, by Lenglart's inequality,

$$\sum \int r((a_{niv}/a_{ni})^{1/2}-1)1\{|(a_{niv}/a_{ni})^{1/2}-1| \leqslant \varepsilon_n\}\, \mathrm{d}X_{ni}(s)$$

$$\leqslant 2\varepsilon_n \sum \int ((a_{niv}/a_{ni})^{1/2}-1)^2\, \mathrm{d}X_{ni}(s) = \mathrm{o}_{P_n}(1).$$

(iii) To show that

$$\sum \int ((a_{niv}/a_{ni})^{1/2}-1)^2 1\{|(a_{niv}/a_{ni})^{1/2}-1| > \varepsilon\}\, \mathrm{d}M_{ni}(s) = \mathrm{o}_{P_n}(1), \tag{2.11}$$

replace $\mathrm{d}M_{ni}(s)$ by $\mathrm{d}X_{ni}(s) - a_{ni}\, \mathrm{d}s$ and apply the argument of part (i) of the proof to each integral.

(iv) Since (2.11) is of order $\mathrm{o}_{P_n}(1)$ for each $\varepsilon > 0$, the same is true for some sequence $\varepsilon_n \to 0$. To show that the second term of the remainder (2.9) is of order $\mathrm{o}_{P_n}(1)$, it remains to prove

$$\sum \int ((a_{niv}/a_{ni})^{1/2}-1)^2 1\{|(a_{niv}/a_{ni})^{1/2}-1| \leqslant \varepsilon_n\}\, \mathrm{d}M_{ni}(s) = \mathrm{o}_{P_n}(1).$$

This follows from

$$\left\langle \sum \int ((a_{niv}/a_{ni})^{1/2}-1)^2 1\{|(a_{niv}/a_{ni})^{1/2}-1| \leqslant \varepsilon_n\}\, \mathrm{d}M_{ni}(s) \right\rangle$$

$$= \sum \int ((a_{niv}/a_{ni})^{1/2}-1)^4 1\{|(a_{niv}/a_{ni})^{1/2}-1| \leqslant \varepsilon_n\} a_{ni}\, \mathrm{d}s$$

$$\leqslant \varepsilon_n^2 \sum \int ((a_{niv}/a_{ni})^{1/2}-1)^2 a_{ni}\, \mathrm{d}s = \mathrm{o}_{P_n}(1). \qquad \square$$

Define the measure Q_{nv} by

$$\mathrm{d}Q_{nv} = \exp(L_{nv})\, \mathrm{d}P_n. \tag{2.12}$$

We call $\{Q_{nv}: v \in V\}$ the *local model.*

We call a functional $k: \Theta \to \mathbb{R}$ *differentiable* at ϑ with *gradient* $g \in V$ if

$$c_n(k(\vartheta_{nv}) - k(\vartheta)) \to (v, g) \quad \text{for } v \in V. \tag{2.13}$$

We call an estimator-sequence $(\hat{k}_n)$ *regular* for k at ϑ with *limit* R if

$$Q_{n,rg} \circ c_n(\hat{k}_n - k(\vartheta_{n,rg})) \Rightarrow R \quad \text{for } r \in \mathbb{R}. \tag{2.14}$$

2.15. Convolution Theorem. *Let $k: \Theta \to \mathbb{R}$ be differentiable at ϑ with gradient $g \in V$. Let $(\hat{k}_n)$ be regular for k at ϑ. Then there exists a distribution S such that*

$$P_n \circ (Z_n(g), c_n(\hat{k}_n - k(\vartheta)) - Z_n(g)) \Rightarrow \mathrm{N}(0, \|g\|^2) \times S.$$

Proof. We reduce the problem to estimating the parameter in a *one-dimensional local* model. Write

$$T_n = c_n(\hat{k}_n - k(\vartheta)).$$

Differentiability of k lets us replace $c_n(k(\vartheta_{n,rg}) - k(\vartheta))$ by $r(g, g) = r\|g\|^2$. Similarly, regularity of $\hat{k}_n$ may be written as

$$Q_{n,rg} \circ (T_n - r\|g\|^2) \Rightarrow R.$$

The assertion we wish to prove is then

$$P_n \circ (Z_n(g), T_n - Z_n(g)) \Rightarrow \mathrm{N}(0, \|g\|^2) \times S. \tag{2.16}$$

Consider the one-dimensional model

$$Q_{n,rg/\|g\|^2}, \quad r \in \mathbb{R}$$

(called the *least favorable submodel*). It is parametrized by the values of the linear functional $(\cdot, g)$:

$$(rg/\|g\|^2, g) = r.$$

The sequence of distributions of T_n is *regular* in the sense that it is asymptotically shift equivariant:

$$Q_{n,rg/\|g\|^2} \circ (T_n - r) \Rightarrow R.$$

Note that R is a probability measure by (2.14) with $r = 0$. Local asymptotic normality of $Q_{n,rg}$ in the sense of Proposition 2.5 implies that $Q_{n,rg}$ and P_n are contiguous, although $Q_{n,rg}$ is not necessarily a probability measure. Assertion (2.16) now follows by adapting Bickel's proof of Hájek's convolution theorem, as in Droste and Wefelmeyer (1984, p. 135, Theorem 2.3). □

The theorem implies that

$$P_n \circ c_n(\hat{k}_n - k(\vartheta)) \Rightarrow \mathrm{N}(0, \|g\|^2) \times S.$$

The Hájek–LeCam convolution theorem is stated in this form; see e.g. LeCam (1986, p. 128, Proposition 2). This justifies calling an estimator-sequence $(\hat{k}_n)$ *efficient* for k at ϑ if

$$P_n \circ c_n(\hat{k}_n - k(\vartheta)) \Rightarrow \mathrm{N}(0, \|g\|^2). \tag{2.17}$$

Other efficiency concepts based on the partially specified likelihood have been introduced by Jacod (1990b, Section 3). These refer to general semimartingales X_n, with characteristics depending on a *finite-dimensional* parameter. One is a Cramér–Rao bound. The other two are variants of the optimality criteria for estimating functions discussed in Godambe and Heyde (1987). For the case $\mathbb{G}_n = \mathbb{F}_n$, these variants have also been introduced by Sørensen (1990, Theorem 3.2). For X_n a counting process, Greenwood and Wefelmeyer (1989c) give conditions under which the optimality criteria are compatible with the asymptotic efficiency concept introduced above.

The relation between efficiency based on the partially specified likelihood and efficiency in a fully specified model is discussed in Greenwood and Wefelmeyer (1989d). For a wide class of full specifications of (X_n, Y_n) as a semimartingale, the likelihood contains the partially specified likelihood as a factor, while the other factor does not depend on ϑ. Then the partially specified likelihood is a true likelihood in every submodel in which the characteristics of Y_n are *fixed.* Furthermore, regularity (2.14) of an estimator-sequence refers to a submodel which contains the probability measure P_n chosen in $\mathscr{P}_{n\vartheta}$. Finally, the asymptotic variance bounds in the partially and fully specified models coincide.

Let $(\hat{k}_n)$ be efficient for k at ϑ in the sense of (2.17). Then the distribution S in Theorem 2.15 equals the Dirac measure at 0. Hence Theorem 2.15 implies a *stochastic* approximation of the estimator-sequence:

$$c_n(\hat{k}_n - k(\vartheta)) = Z_n(g) + o_{P_n}(1). \tag{2.18}$$

This result is already contained in LeCam (1953, p. 327, Theorem 14) and Hájek (1972, p. 186, Theorem 4.1). See also LeCam (1986, p. 115, Theorem 1) or Strasser (1985, p. 315, Theorem 63.6).

2.19. Remark. The Cramér–Wold theorem and LeCam's third lemma imply that an estimator-sequence with (2.18) fulfills, for *all* $v \in V$,

$$Q_{nv} \circ c_n(\hat{k}_n - k(\vartheta_{nv})) \Rightarrow \mathrm{N}(0, \|g\|^2).$$

This means that the estimator-sequence is regular in a stronger sense than (2.14).

2.20. Remark. Jacod (1987, p. 68, Theorem (6.3); 1990a, Théorème 7.12) proves (for general semimartingales) that the distribution of the partially specified log-likelihood L_{nv} under P_n is asymptotically normal. Hence, by contiguity, the sequence of experiments $\{Q_{nv}: v \in V\}$ converges weakly to a Gaussian shift experiment. (See e.g. LeCam, 1986, p. 173, Lemma 1.) If V is separable, this implies the *existence* of a linear process Z_n such that the partially specified log-likelihood admits a stochastic approximation as in (2.6):

$$L_{nv} = Z_n(v) - \tfrac{1}{2}\|v\|^2 + o_{P_n}(1).$$

(See LeCam, 1986, p. 176, Proposition 1.) The explicit form (2.7) of Z_n obtained in Proposition 2.5, besides being of interest in itself, is occasionally useful. For example, efficiency of an estimator $\hat{k}_n$ for a functional k can be checked by verifying the stochastic approximation (2.18) instead of determining the asymptotic distribution of the estimator. The explicit form of Z_n can also be used to define efficient estimating equations; see Greenwood and Wefelmeyer (1989b).

We will see in Corollary 2.22 below that the Convolution Theorem 2.15 for one-dimensional functionals implies a convolution theorem for finite-dimensional functionals $k: \Theta \to \mathbb{R}^p$.

We call $k=(k_1,\dots,k_p)'$ *differentiable* at ϑ with *gradient* $g=(g_1,\dots,g_p)'$ if the components k_j are differentiable in the sense of (2.13) with gradients g_j.

We call an estimator-sequence $\hat{k}_n=(\hat{k}_{n1},\dots,\hat{k}_{np})'$ *regular* for k at ϑ with *limit* R if, for v in the linear span of $g_1,\dots,g_p$,

$$Q_{nv}\circ c_n(\hat{k}_n-k(\vartheta_{nv}))\Rightarrow R. \tag{2.21}$$

2.22. Corollary. *Let $k:\Theta\to\mathbb{R}^p$ be differentiable at ϑ with gradient $g\in V^p$. Let $(\hat{k}_n)$ be regular for k at ϑ with limit R. Then there exists a distribution S such that*

$$P_n\circ(Z_n(g),\, c_n(\hat{k}_n-k(\vartheta))-Z_n(g))\Rightarrow \mathrm{N}(0,(g,g'))\times S.$$

Here (g,g') denotes the matrix with entries the acuity inner product (g_i,g_j).

Proof. Fix $b\in\mathbb{R}^p$. Define $q(r)=b'r$ for $r\in\mathbb{R}^p$. Then $b'k$ is differentiable at ϑ with gradient $b'g$. Furthermore, $b'\hat{k}_n$ is regular for $b'k$ at ϑ with limit $R\circ q$. By the Convolution Theorem 2.15 there exists a distribution S_b on $\mathbb{R}$ such that

$$P_n\circ(b'Z_n(g),\, c_nb'(\hat{k}_n-k(\vartheta))-b'Z_n(g))\Rightarrow \mathrm{N}(0,b'(g,g')b)\times S_b.$$

Since $c_n(\hat{k}_n-k(\vartheta))$ and $Z_n(g)$ are tight, there exists a distribution S on $\mathbb{R}^p$ such that for some subsequence,

$$P_n\circ(c_n(\hat{k}_n-k(\vartheta))-Z_n(g))\Rightarrow S.$$

Hence $S_b=S\circ q$ and

$$\mathrm{N}(0,b'(g,g')b)\times S_b=(\mathrm{N}(0,(g,g'))\times S)\circ q.$$

In particular, $R\circ q=(\mathrm{N}(0,(g,g'))\times S)\circ q$. Hence S is independent of the subsequence. The assertion now follows from the Cramér–Wold theorem. □

3. The additive risk model

The additive risk model, introduced by Aalen (1980), has an intensity process of the form

$$a_{ni\vartheta}(s)=Y_{ni}(s)\lambda(s),\quad s\in[0,1]. \tag{3.1}$$

Here Y_{ni} is a p-dimensional row vector of *covariate processes,* and λ is a p-dimensional column vector of bounded *hazard functions.*

Set $\vartheta=\lambda$. Fix $P_n\in\mathcal{P}_{n\vartheta}$. Introduce weights

$$W_{ni}=(Y_{ni}\lambda)^{-1}.$$

Define

$$W_n=\operatorname{diag} W_{ni},\qquad \Lambda=\operatorname{diag}\lambda_j.$$

Let Y_n denote the matrix with rows Y_{ni}. Assume that there exists a matrix function U with largest and smallest eigenvalue bounded and bounded away from 0, respectively, such that

$$\sup_{t\in[0,1]} |c_n^{-2} Y_n' W_n Y_n - U| \to 0 \quad (P_n). \tag{3.2}$$

We introduce a *local model* as follows. Let V be the set of all *bounded* measurable functions from $[0,1]$ into $\mathbb{R}^p$. For each *local parameter* $v = (v_1, \ldots, v_p)' \in V$ and n sufficiently large define

$$\lambda_{nvj} = (1 + c_n^{-1} v_j)\lambda_j. \tag{3.3}$$

Set $a_{niv} = Y_{ni}\lambda_{nv}$. Let us check that the intensity process (3.1) fulfills conditions (2.1′), (2.2) and (2.3).

(1) *Hellinger differentiability.* From definition (3.3),

$$a_{niv}/a_{ni} = Y_{ni}\lambda_{nv}(Y_{ni}\lambda)^{-1} = Y_{ni}\lambda_{nv}W_{ni} = 1 + c_n^{-1} v'\Lambda Y_{ni} W_{ni}.$$

Hence (2.1′) holds with $D_{ni}(v) = v'\Lambda Y_{ni} W_{ni}$.

(2) *Lindeberg condition.* Let $|v_j| \le c$ for $j = 1, \ldots, p$. Then

$$|D_{ni}(v)| = |v'\Lambda Y_{ni} W_{ni}| \le c Y_{ni}\lambda W_{ni} = c.$$

This implies (2.2).

(3) *Nonrandom Hellinger limit.* Relation (3.2) implies

$$c_n^{-2} \sum \int D_{ni}(v)^2 a_{ni}\, \mathrm{d}s = c_n^{-2} \int v'\Lambda Y_n' W_n Y_n \Lambda v\, \mathrm{d}s$$
$$= \int v'\Lambda U \Lambda v\, \mathrm{d}s + o_{P_n}(1).$$

Hence (2.3) holds with

$$\|v\|^2 = \int v'\Lambda U \Lambda v\, \mathrm{d}s. \tag{3.4}$$

Proposition 2.5 now implies that the partially specified log-likelihood is locally asymptotically normal,

$$L_{nv} = c_n^{-1} \sum_{i=1}^{n} \int v'\Lambda Y_{ni} W_{ni}\, \mathrm{d}M_{ni\vartheta}(s) - \tfrac{1}{2} \int v'\Lambda U \Lambda v\, \mathrm{d}s + o_{P_n}(1).$$

Here $M_{ni\vartheta}(t) = X_{ni}(t) - \int_0^t Y_{ni}\lambda\, \mathrm{d}s$. By (3.4), the acuity is

$$(v, w) = \int v'\Lambda U \Lambda w\, \mathrm{d}s. \tag{3.5}$$

Now we can determine the asymptotic covariance bound for estimators of the vector of cumulative hazards $\int_0^t \lambda\, \mathrm{d}s$.

Writing the cumulative hazards as a *row* vector for convenience, we introduce the p-dimensional functional

$$k(\lambda)=\int_0^t \lambda'\,\mathrm{d}s.$$

From (3.3) we obtain for $v\in V$ and $j=1,\ldots,p$,

$$c_n\left(\int_0^t \lambda_{nvj}\,\mathrm{d}s-\int_0^t \lambda_j\,\mathrm{d}s\right)=\int_0^t v_j\lambda_j\,\mathrm{d}s.$$

Hence

$$c_n(k(\lambda_{nv})-k(\lambda))=\int_0^t v'\Lambda\,\mathrm{d}s.$$

The right-hand side is a linear functional of v. A gradient, g, of the functional k in the sense of (2.13) is determined by expressing this linear functional in terms of the acuity:

$$\int_0^t v'\Lambda\,\mathrm{d}s=(v,g)\quad\text{for } v\in V. \tag{3.6}$$

Note that g is a $p\times p$ matrix such that the j-th column is the gradient of the j-th component k_j of the functional k.

Using the explicit form (3.5) of the acuity, we can write equation (3.6) as

$$\int_0^t v'\Lambda\,\mathrm{d}s=\int v'\Lambda U\Lambda g\,\mathrm{d}s. \tag{3.7}$$

This is solved by

$$g(s)=\Lambda(s)^{-1}U(s)^{-1}1_{[0,t]}(s),\quad s\in[0,1].$$

The function g is bounded and therefore in V. Hence the assumptions of the Convolution Theorem 2.15 are fulfilled, and the covariance bound is obtained by applying (3.7) for $v=g$:

$$(g',g)=\int_0^t g'\Lambda\,\mathrm{d}s=\int_0^t U^{-1}\Lambda^{-1}\Lambda\,\mathrm{d}s=\int_0^t U^{-1}\,\mathrm{d}s.$$

We note in passing that the covariance bound can also be obtained by calculating the acuity (3.5) for $v=g'$, $w=g$:

$$(g',g)=\int g'\Lambda U\Lambda g\,\mathrm{d}s=\int_0^t U^{-1}\Lambda^{-1}\Lambda U\Lambda\Lambda^{-1}U^{-1}\,\mathrm{d}s=\int_0^t U^{-1}\,\mathrm{d}s.$$

McKeague (1988b) shows that this is the asymptotic covariance of the weighted least squares estimator introduced by Huffer and McKeague (1987). Hence this estimator is efficient. It is easy to check that the estimator admits a stochastic approximation (2.18) and is therefore regular by Remark 2.19; see Greenwood and Wefelmeyer (1989b, Section 7).

4. The proportional hazards model

The counting process version of the by now classical proportional hazards model of Cox (1972) was introduced by Andersen and Gill (1982). Its intensity process is of the form

$$a_{ni\vartheta}(s) = C_{ni}(s)\lambda(s)\exp(\beta' Y_{ni}(s)), \quad s \in [0, 1]. \tag{4.1}$$

Here C_{ni} is a *censoring process* taking only values 0 and 1, Y_{ni} is a p-dimensional vector of *covariate processes*, β is a p-dimensional vector of *regression coefficients*, and λ is a bounded *baseline hazard function*. Assume for simplicity that the covariate processes Y_{ni} are uniformly bounded.

Set $\vartheta = (\beta, \lambda)$. Fix $P_n \in \mathscr{P}_{n\vartheta}$. Our regularity conditions are similar to those of Andersen and Gill (1982). Let $n^{-1/4}c_n \to \infty$. Assume that there exist bounded scalar, vector and matrix functions S_0, S_1, S_2, with S_0 bounded away from 0, such that

$$\sup_{t\in[0,1]} \left| c_n^{-2} \sum_{i=1}^{n} C_{ni}\exp(\beta' Y_{ni}) - S_0 \right| \to 0 \quad (P_n), \tag{4.2}$$

$$\sup_{t\in[0,1]} \left| c_n^{-2} \sum_{i=1}^{n} C_{ni} Y_{ni}\exp(\beta' Y_{ni}) - S_1 \right| \to 0 \quad (P_n), \tag{4.3}$$

$$\sup_{t\in[0,1]} \left| c_n^{-2} \sum_{i=1}^{n} C_{ni} Y_{ni} Y_{ni}'\exp(\beta' Y_{ni}) - S_2 \right| \to 0 \quad (P_n). \tag{4.4}$$

Assume that

$$\Sigma = \int (S_2 - S_0^{-1} S_1 S_1')\lambda \, \mathrm{d}s \text{ is nonsingular.} \tag{4.5}$$

We introduce a *local model* at $\vartheta = (\beta, \lambda)$ as follows. Let B be the set of all *bounded* measurable functions on [0, 1]. Set $V = \mathbb{R}^p \times B$. For each *local parameter* $(b, v) \in V$ define

$$\beta_{nb} = \beta + c_n^{-1} b, \tag{4.6}$$

$$\lambda_{nv} = (1 + c_n^{-1} v)\lambda. \tag{4.7}$$

Set $a_{nibv} = C_{ni}\lambda_{nv}\exp(\beta_{nb}' Y_{ni})$. Let us check that the intensity process (4.1) fulfills conditions (2.1′), (2.2) and (2.3).

(1) *Hellinger differentiability.* Set $r(x) = \exp(x) - 1 - x$. From definitions (4.6) and (4.7), when $C_{ni} = 1$,

$$a_{nibv}/a_{ni} - 1 = c_n^{-1}(b' Y_{ni} + v) + c_n^{-2} v b' Y_{ni} + (1 + c_n^{-1} v) r(c_n^{-1} b' Y_{ni}).$$

Since the covariates are uniformly bounded, we have

$$c_n^{-4} \sum \int v^2 a_{ni} \, \mathrm{d}s = o_{P_n}(1),$$

and condition (2.1′) holds with

$$D_{ni}(b, v) = b'Y_{ni} + v.$$

(2) *Lindeberg condition.* Since $D_{ni}(b, v)$ is bounded, (2.2) holds.

(3) *Nonrandom Hellinger limit.* By (4.2) to (4.4),

$$c_n^{-2} \sum \int D_{ni}(b, v)^2 a_{ni} \, ds = c_n^{-2} \sum \int (b'Y_{ni} + v)^2 C_{ni} \lambda \exp(\beta' Y_{ni}) \, ds$$

$$\to \int (b'S_2 b + 2b'S_1 v + S_0 v^2)\lambda \, ds \quad (P_n).$$

Hence (2.3) holds with

$$\|(b, v)\|^2 = \int (b'S_2 b + 2b'S_1 v + S_0 v^2)\lambda \, ds. \tag{4.8}$$

Proposition 2.5 now implies that the partially specified log-likelihood is locally asymptotically normal:

$$L_{nbv} = n^{-1/2} \sum_{i=1}^{n} \int (b'Y_{ni} + v) \, dM_{ni\vartheta}(s)$$
$$- \tfrac{1}{2} \int (b'S_2 b + 2b'S_1 v + S_0 v^2)\lambda \, ds + o_{P_n}(1).$$

Here

$$M_{ni\vartheta}(t) = X_{ni}(t) - \int_0^t C_{ni} \lambda \exp(\beta' Y_{ni}) \, ds.$$

By (4.8) the acuity is

$$((b, v), (c, w)) = \int (b'S_2 c + b'S_1 w + vS_1' c + vS_0 w)\lambda \, ds. \tag{4.9}$$

We are now ready to determine the asymptotic covariance bounds for estimators of the regression coefficients and the cumulative baseline hazard.

Let us first consider the vector β of *regression coefficients.* Writing β as a *row* vector for convenience, we introduce the p-dimensional functional

$$k(\beta, \lambda) = \beta'.$$

By definition of β_{nb} we have

$$c_n(k(\beta_{nb}, \lambda_{nv}) - k(\beta, \lambda)) = c_n(\beta'_{nb} - \beta') = b'.$$

The right-hand side can be considered as a linear functional on V. A gradient, say (b_β, v_β), of the functional k in the sense of (2.13) is determined by expressing this linear functional in terms of the acuity:

$$b' = ((b, v), (b_\beta, v_\beta)) \quad \text{for } (b, v) \in V. \tag{4.10}$$

Note that (b_β, v_β) is a row of elements $(b_{\beta_j}, v_{\beta_j})$ in $\mathbb{R}^p \times B$ which are gradients of the j-th component of the functional k.

Using the explicit form (4.9) of the acuity, we obtain the two equations

$$b' = b' \int (S_2 b_\beta + S_1 v_\beta)\lambda \, \mathrm{d}s \quad \text{for } b \in \mathbb{R}^p,$$

$$0 = \int (v S_1' b_\beta + v S_0 v_\beta)\lambda \, \mathrm{d}s \quad \text{for } v \in B.$$

The second equation gives $0 = S_1' b_\beta + S_0 v_\beta$, i.e.

$$v_\beta = -S_0^{-1} S_1' b_\beta.$$

Inserting into the first equation, we obtain $b_\beta = \Sigma^{-1}$ with Σ defined in (4.5). Hence

$$v_\beta = -S_0^{-1} S_1' \Sigma^{-1},$$

and the gradient is

$$(b_\beta, v_\beta) = (\Sigma^{-1}, -S_0^{-1} S_1' \Sigma^{-1}).$$

(This can be checked by showing that (4.10) holds.) Since $S_0^{-1} S_1$ is bounded, the gradient is in V. Hence all assumptions of the Convolution Theorem 2.15 are fulfilled. The covariance bound for estimators of β is now obtained by applying (4.10) for $(b, v) = (b_\beta, v_\beta)$:

$$((b_\beta, v_\beta)', (b_\beta, v_\beta)) = b_\beta' = \Sigma^{-1}.$$

(The same result is obtained by applying (4.9) for $(b, v) = (b_\beta, v_\beta)'$, $(c, w) = (b_\beta, v_\beta)$; the calculation is, however, slightly longer.)

Andersen and Gill (1982, p. 1106, Theorem 3.2) have shown that Σ^{-1} is the asymptotic covariance of the Cox estimator. Hence this estimator is efficient. According to Remark 2.19, the estimator is regular if it admits a stochastic approximation (2.18). Such a stochastic approximation may be obtained from Andersen and Gill (1982, p. 1103) or from Greenwood and Wefelmeyer (1989b, Section 8).

The covariance bound has already been obtained in several special cases (with one-dimensional covariates). For a fully specified model see Dzhaparidze (1985). For a partially specified model, Dzhaparidze (1987) states that the Cox estimator is efficient in a class of asymptotically linear estimators. Different methods for calculating the bound may be found in Begun et al. (1983, p. 448, Example 4), Begun and Wellner (1983, for the two-sample case), Ritov and Wellner (1988, p. 205, Section 4), and Klaassen (1988, Section 3). These refer to the classical version of the Cox model, with time-independent covariates.

Let us now turn to determining the asymptotic covariance bound for estimators of the *cumulative baseline hazard* $\int_0^t \lambda \, \mathrm{d}s$. We introduce the one-dimensional functional

$$k_t(\beta, \lambda) = \int_0^t \lambda \, \mathrm{d}s.$$

From (4.7) we obtain for $(b, v) \in V$,

$$c_n(k_t(\beta_{nb}, \lambda_{nv}) - k_t(\beta, \lambda)) = c_n\left(\int_0^t \lambda_{nv} \, \mathrm{d}s - \int_0^t \lambda \, \mathrm{d}s\right) = \int_0^t v\lambda \, \mathrm{d}s.$$

A gradient, say (b_t, v_t), of the functional k_t in the sense of (2.13) is determined by expressing the linear functional on the right-hand side in terms of the acuity:

$$\int_0^t v\lambda \, ds = ((b, v), (b_t, v_t)) \quad \text{for all } (b, v) \in V. \tag{4.11}$$

Using the explicit form (4.9) of the acuity, we obtain the two equations

$$0 = \int (S_2 b_t + S_1 v_t)\lambda \, ds,$$

$$\int_0^t v\lambda \, ds = \int (vS_1' b_t + vS_0 v_t)\lambda \, ds \quad \text{for } v \in B.$$

The second equation gives $1_{[0,t]} = S_1' b_t + S_0 v_t$, i.e.

$$v_t = -S_0^{-1} S_1' b_t + S_0^{-1} 1_{[0,t]}.$$

Replacing v_t in the first equation by this expression, we obtain $0 = \Sigma b_t + c_t$ with Σ as in (4.5) and

$$c_t = \int_0^t S_0^{-1} S_1 \lambda \, ds.$$

Hence $b_t = -\Sigma^{-1} c_t$, and the gradient is

$$(b_t, v_t) = (-\Sigma^{-1} c_t, S_0^{-1} S_1' \Sigma^{-1} c_t + S_0^{-1} 1_{[0,t]}).$$

Note that v_t is bounded. Hence (b_t, v_t) is in V, and all assumptions of the Convolution Theorem 2.15 are fulfilled. The variance bound for estimators of $\int_0^t \lambda \, ds$ is now obtained by applying (4.11) for $(b, v) = (b_t, v_t)$:

$$((b_t, v_t), (b_t, v_t)) = \int_0^t v_t \lambda \, ds = c_t' \Sigma^{-1} c_t + \int_0^t S_0^{-1} \lambda \, ds.$$

It follows from Andersen and Gill (1982, p. 1108, Theorem 3.4) that this is the asymptotic variance of their Breslow-type estimator. Hence this estimator is efficient.

For the classical Cox model, with time-independent covariates, the bound was obtained by Begun et al. (1983, p. 450).

4.12. Remark. Prentice and Self (1983) suggest replacing the exponential function in the Cox model (4.1) by some other *known* function r:

$$a_{ni\vartheta}(s) = C_{ni}(s)\lambda(s) r(\beta' Y_{ni}(s)), \quad s \in [0, 1].$$

This model can be treated in exactly the same way as the Cox model. We do not formulate the assumptions precisely.

Introduce local parameters as in (4.6) and (4.7). Set

$$r^{(1)}(y) = \partial_y r(y), \qquad u = r^{(1)}/r.$$

For $r = \exp$ we have $r^{(1)} = \exp$ and $u \equiv 1$. By Taylor expansion,

$$r(\beta_{nb}' y) = r(\beta' y)(1 + n^{-1/2} b' y u(b' y) + o(n^{-1/2})).$$

As for the Cox model we obtain (2.1′), with

$$D_{ni}(b, v) = b'Y_{ni}u(b'Y_{ni}) + v.$$

Assume stability conditions

$$c_n^{-2} \sum C_{ni} r(\beta' Y_{ni}) \to S_0,$$
$$c_n^{-2} \sum C_{ni} Y_{ni} r(\beta' Y_{ni}) \to S_1,$$
$$c_n^{-2} \sum C_{ni} Y_{ni} Y'_{ni} r(\beta' Y_{ni}) \to S_2.$$

With these S_j, one can prove local asymptotic normality and calculate gradients and covariance bounds as in the Cox model. Prentice and Self (1983, p. 809, Theorem 2.1, and p. 811, Theorem 2.2) have obtained the asymptotic distributions of a Cox-type estimator for the regression coefficients and a Breslow-type estimator for the cumulative baseline hazard in their model. The results correspond to those of Andersen and Gill (1982). Hence the estimators are efficient.

Acknowledgment

This work was begun while the authors were visiting the Department of Mathematical Sciences at the Johns Hopkins University. We are most grateful to Robert J. Serfling for his kind hospitality. The results were presented at an Oberwolfach meeting on martingale methods in statistics in December, 1988. We wish to thank Richard Gill, Jean Jacod, Ian McKeague, Eric Slud and Helmut Strasser for their comments and access to unpublished manuscripts. The paper was finished while the second author was visiting the Department of Mathematics at the University of British Columbia, Vancouver. We acknowledge with pleasure the numerous suggestions and corrections made by the referees.

References

O.O. Aalen, A model for nonparametric regression analysis of counting processes, in: W. Klonecki, A. Kozek and J. Rosiński, eds., Mathematical Statistics and Probability Theory, Lecture Notes in Statistics No. 2 (Springer, New York, 1980) 1-25.

P.K. Andersen, Ø. Borgan, R.D. Gill and N. Keiding, Statistical Methods for Counting Processes, Book manuscript (1990).

P.K. Andersen and R.D. Gill, Cox's regression model for counting processes: a large sample study, Ann. Statist. 10 (1982) 1100-1120.

J.M. Begun, W.J. Hall, W.-M. Huang and J.A. Wellner, Information and asymptotic efficiency in parametric-nonparametric models, Ann. Statist. 11 (1983) 432-452.

J.M. Begun and J.A. Wellner, Asymptotic efficiency of relative risk estimates, in: P.K. Sen, ed., Contributions to Statistics: Essays in Honour of Norman Lloyd Johnson (North-Holland, Amsterdam, 1983) 47-62.

N. Breslow, Discussion on Professor Cox's paper, J. Roy. Statist. Soc. Ser. B 34 (1972) 216-217.

N. Breslow, Covariance analysis of censored survival data, Biometrics 30 (1974) 89-99.

D.R. Cox, Regression models and life tables (with discussion), J. Roy. Statist. Soc. Ser. B 34 (1972) 187-220.

D.R. Cox, Partial likelihood, Biometrika 62 (1975) 69–76.
W. Droste and W. Wefelmeyer, On Hájek's convolution theorem, Statist. and Decisions 2 (1984) 131–144.
K. Dzhaparidze, On asymptotic inference about intensity parameters of a counting process, Bull. Int. Statist. Inst. 51 (1985) Vol. 4, 23.2-1—23.2-15.
K. Dzhaparidze, On asymptotic efficiency of the Cox estimator, in: Yu. Prohorov and V.V. Sazonov, eds, Proceedings of the 1st World Congress of the Bernoulli Society Vol. 2, Mathematical Statistics. Theory and Applications (VNU Science Press, Utrecht, 1987) 59–61.
R.D. Gill, Notes on product integration, likelihood and partial likelihood for counting processes, non-informative and independent censoring, Unpublished manuscript (1985).
V.P. Godambe and C.C. Heyde, Quasi-likelihood and optimal estimation, Int. Statist. Rev. 55 (1987) 231–244.
P.E. Greenwood, Partially specified semimartingale experiments, in: N.U. Prabhu, ed., Statistical Inference from Stochastic Processes, Contemporary Math. Vol. 80 (Amer. Math. Soc., Providence, RI, 1988) 1–17.
P.E. Greenwood and W. Wefelmeyer, Efficiency bounds for estimating functionals of stochastic processes, Preprints in Statistics 117, Dept. of Mathematics, University of Cologne (1989a).
P.E. Greenwood and W. Wefelmeyer, Efficient estimating equations for nonparametric filtered models, in: I.V. Basawa and N.U. Prabhu, eds., Statistical Inference in Stochastic Processes, Vol. 1 (1989b, to appear).
P.E. Greenwood and W. Wefelmeyer, On optimal estimating functions for partially specified counting process models, Preprints in Statistics 123, Dept. of Mathematics, University of Cologne (1989c).
P.E. Greenwood and W. Wefelmeyer, Partially and fully specified semimartingale models and efficiency, Preprints in Statistics 124, Dept. of Mathematics, University of Cologne (1989d).
J. Hájek, Local asymptotic minimax and admissibility in estimation, Proc. Sixth Berkeley Symp. Math. Statist. Probab. Vol. 1 (1972) 175–194.
F.W. Huffer and I.W. McKeague, Survival analysis using additive risk models, Tech. Rep. No. 396, Dept. of Statistics, Stanford University (1987).
J. Jacod, Multivariate point processes: predictable projection, Radon–Nikodym derivatives, representation of martingales, Z. Wahrsch. Verw. Gebiete 31 (1975) 235–253.
J. Jacod, Partial likelihood process and asymptotic normality, Stochastic Process. Appl. 26 (1987) 47–71.
J. Jacod, Sur le processus de vraisemblance partielle, Ann. Inst. H. Poincaré Probab. Statist. 26 (1990a) 299–329.
J. Jacod, Regularity, partial regularity, partial information process for a filtered statistical model, Theory Probab. Related Fields (1990b, to appear).
J. Jacod and A.N. Shiryaev, Limit Theorems for Stochastic Processes, Grundlehren der mathematischen Wissenschaften Vol. 288 (Springer, Berlin, 1987).
C.A.J. Klaassen, Efficient estimation in the Cox model for survival data, Unpublished manuscript (1988).
L. LeCam, On some asymptotic properties of maximum likelihood estimates and related Bayes estimates, Univ. Calif. Publ. Statist. 1 (1953) 277–330.
L. LeCam, Asymptotic Methods in Statistical Decision Theory, Springer Series in Statistics (Springer, New York, 1986).
I.W. McKeague, A counting process approach to the regression analysis of grouped survival data, Stochastic Process. Appl. 28 (1988a) 221–239.
I.W. McKeague, Asymptotic theory of weighted least squares estimators in Aalen's additive risk model, in: N.U. Prabhu, ed., Statistical Inference from Stochastic Processes, Contemporary Math. Vol. 80 (Amer. Math. Soc., Providence, RI, 1988b) 139–152.
I.W. McKeague and K.J. Utikal, Stochastic calculus as a tool in survival analysis: a review, Appl. Math. Comput. (1990, to appear).
R.L. Prentice and S.G. Self, Asymptotic distribution theory for Cox-type regression models with general relative risk form, Ann. Statist. 11 (1983) 804–813.
Y. Ritov and J.A. Wellner, Censoring, martingales and the Cox model, in: N.U. Prabhu, ed., Statistical Inference from Stochastic Processes, Contemporary Math. Vol. 80 (Amer. Math. Soc., Providence, RI, 1988) 191–219.
E. Slud, Martingale Methods in Statistics, Book manuscript (1986).
M. Sørensen, On quasi likelihood for semimartingales, Stochastic Process. Appl. 35 (1990) 331–346.
H. Strasser, Mathematical Theory of Statistics (De Gruyter, Berlin, 1985).